

Optimizing Ontology Alignment for Nuclear Information System

M.Ramesh

PG Scholar, Software Engineering, Indian Institute of Information Technology, Srirangam,
Tiruchirappalli, Tamilnadu, India.

P.Karthikeyan

Assistant Professor, Department of CSE/IT, BIT Campus, Anna University,
Tiruchirappalli, Tamilnadu, India.

Dr.N.Madurai Meenachi

Scientific Officer, Resource Management Group, Indira Gandhi Centre for Atomic Research,
Kalpakkam-603102, Tamilnadu, India.

Dr. M.Sai Baba

Associate Director, Resource Management Group, Indira Gandhi Centre for Atomic Research,
Kalpakkam-603102, Tamilnadu, India.

Abstract – Ontology is the interrelationship of entities which is used for searching hierarchical data. It is the foundation of the semantic web that plays an important role in enabling interoperability between applications. Ontology mapping is an essential part of the semantic web which have role of finding alignments among the entities of the given ontologies. Ontology mapping is a necessary step for establishing interoperation and knowledge sharing among semantic web applications. The proposed methodology for mapping the ontologies gives more efficiency than existing methodologies. It classifies the given input into classes, subclass and properties using OWLAPI. It matches using S-Match technique in memetic algorithm to get optimized results. As shown by experiments, this approach gives best results than existing methods.

Index Terms – ontology alignment; memetic algorithm; smoa distance; heterogeneity problem.

1. INTRODUCTION

The general definition of ontology is an interrelationship of entities which is representation of entities along with the relation and properties. It used for searching hierarchical type of data. Ontology plays an important role in enabling the interoperability between applications [1]. This ontology defines same entity with different names which causes heterogeneity problem for semantic interoperability between applications. To address this problem, ontology alignment plays an important role for increasing and utilization of semantic web based applications. But it is difficult to align ontologies when the size is considerably large [2]. Hence, various ontology alignment systems have been created in recent years. Each system uses different methodologies for

calculating similarity between entities from ontologies. Most of the similarity measure could not provide best results independently, so different types of similarity measures are together aggregated to align their results. This process is called meta-matching which can also be viewed as an optimization problem.

This optimization problem is addressed by many approaches like Evolutionary Approach [3]. This system has been implemented to solve interoperability among heterogeneous system. This ontology alignment process is performed by calculating similarity values among the concepts of different ontologies in terms of lexical, linguistic and structural similarity approaches [4]. The existing genetic algorithm has a problem of knowledge domain. Hence this paper tries to solve this issue using memetic algorithm [5] by different similarity matching technique. This algorithm extends the genetic algorithm with the local optimization methods (e.g., stochastic hill climbing and edge count method) to improve the quality of the solution of problems. The main purpose of this ontology alignment is to solve semantic heterogeneity problem which is main obstacles of semantic interoperability. The system implementation section describes how our proposal calculates ontology alignment in more suitable and in efficient way than genetic algorithm.

2. RELATED WORK

Recently, the ontology alignment problem has been risen and its complexity to develop automatic and semi-automatic ontology alignment systems [1]. The first one among these systems is PROMPT [10]. It is a semi-automatic matching

system which gives a set of suggestions to be used for matching classes and properties in mapping a file. The drawback of this system is it does not produce its output as alignment. The Automatic Ontology matching system is Cupid [11], it calculates the mapping between two ontologies by selecting pair of entities with an aggregated similarity value greater than threshold. This similarity is computed with weights and threshold manually. Since then, various types of automatic ontology alignment system have been implemented by using different techniques. Among these systems, recent systems are MapPSO [12], ASMOV [13], CODI [14], COMA++ [15] and so on.

The memetic aggregation of similarity measures [16] tries to solve semantic heterogeneity problem for different ontologies with same information. To get optimized results, the author used different similarity measures. This method will aggregate the similarity measures without using a prior knowledge about ontologies under reference alignment. It yields good results in terms of alignment quality when related to Ontology Alignment Evaluation Initiative. The No Reference alignment technique was implemented using UIR (Unanimous Improvement Ratio) and MatchFMeasure [17]. This method prevents three defects from the previous methods. It used rough evaluation metric on no reference alignment for appropriate f-measure and UIR. It gives the optimized results with respect to state of the art ontology matching systems.

For matching two large ontologies many techniques can be used such as parallel matching, search space and self-tuning [18]. A cluster based approach [19] will break the matching problem into smaller matching problems. It first uses structure based clustering approach for partition each schema graph into clusters. Then this system efficiently determines the similar type of clusters between every set of clusters which is a small matching task. This technique used to COMA++ to solve the individual matching tasks and then combine their results.

In this Coincidence Based Scoring [20] approach, the given ontologies are interpreted as types of graphs embedded in the metric space in which the coincidence of the structure of the two ontologies is formulated. By using this formulation we can define mechanism to score mappings. The scoring is used to extract good alignment among the number of sets. The three approaches used in this paper are Straight Forward, Genetic Algorithm and Based on Approximative Approach. The Straight forward finds the optimum alignment and checks for all possible alignment so it takes more runtime and also it has limited to small ontologies. But in Genetic Algorithm, it shows some effectiveness for some certain test collections. The final Approximative approach will apply a maximum weight matching method for ontology mapping that would be quite inefficient because it will lose inherent structure and interrelationships of both ontologies. The idea of coincidence

is a measure for evaluating how coincident the two ontologies as a whole.

3. PROPOSED METHOD

The proposed methodology comprises of three steps namely Preprocessing, Optimization module and Alignment module. The first step will preprocess and categorize input files. The second step will do optimization by memetic algorithm. The final step will align the output files with result.

3.1. Preprocessing

The first step of the work is preprocessing of the given input ontology files. The Input ontology file is developed by a knowledge expert using protégé Tool which is open source tool for creating Ontology OWL files. The Input contains Classes, Properties, Instance etc., For every classes there are Superclass, Subclasses for it. Property contains Object Property, Data Property etc., The preprocessing step categorize the each of the classes and properties in Ontology.

After preprocessing step, the given input is converted into small tree structure format like class, its super class and its sub class. All the inputs are converted into this type. It will give the best matching result compared to the existing normal String Matching. For selecting the inputs from the given dataset, we use Memetic algorithm which has few steps for selecting the classes and providing the optimum solution.

3.2. Optimization Module

In this module, Memetic algorithm was used for giving the optimized results in the ontology matching [6]. It gives different types of results for various types of domains. In this paper, nuclear domain is taken as an input domain for which nuclear ontologies are created using protégé tool. This algorithm is the extended version of the GA (Genetic Algorithm) with which includes Stochastic Hill Climbing search for local refining process.

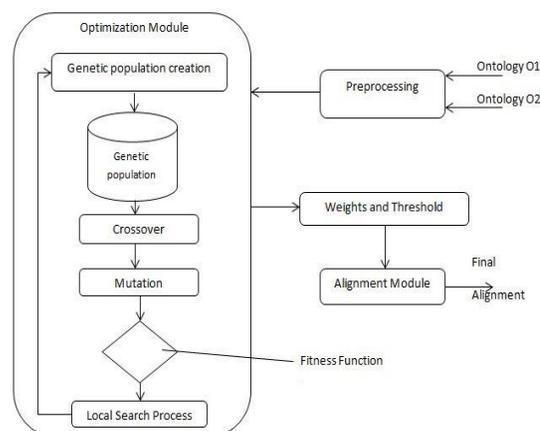


Fig. 1. Architecture of the proposed system.

4. EXPERIMENTAL EVALUATION

4.1. Experimental Setup

The process of creating an initialized population is called as Initialization. The population is usually random generated and also it can be at any size. The created population is called as genetic population. The crossover is like selection also known as Convergence operation. It pulls population towards the local minimum/maximum. The Mutation is also known as Divergence operation. It breaks one or more members of population of the local minimum/maximum.

To implement this system for ontology matching process, we conducted experiments which implemented with Java Server Pages using OWLAPI [8], INIS thesaurus [9] and Java String Similarity Functions. The experiment methodology works as follows. First each classes or entities are extracted from the two given ontologies which are named as concepts. The preprocess step will categorize class with their subclass and properties. Then for each string similarity algorithm in the package, we calculate a similarity score for each pair of classes that belongs to two ontologies respectively. We extract alignment results using memetic algorithm [5]. The experiments were conducted on Intel Core i5 4570 3.20 GHz processors and 4 GB RAM under windows 8. The data set comprises of nuclear ontologies. The inputs are created using the protégé tool with the knowledge of domain expert. We compute the values of precision, recall and F-measure.

The fitness function is used for finding the distance between the two concepts in the ontologies. Here we use SMOA (String Metrics for Ontology Alignment) to calculate distance [7]. It is a well performed method especially made for calculating the similarities of entities in ontology alignment. For example, two given strings $s1$ and $s2$, the SMOA distance between the given string can be defined by following equation:

4.2. Experimental Results

$SMOA(S1,S2)=Comm(S1,S2)- Diff(S1,S2)+ WinkImpr(S1,S2)$
 Where $Comm(S1,S2)$ stands for the Commonality between $S1$ and $S2$, $Diff(S1,S2)$ is the difference and $WinkImpr(S1,S2)$ for improvement of the result.

The local search process uses stochastic hill climbing algorithm. The strategy of this algorithm is to iterate the process of randomly selecting a neighbor for an optimal solution and only accept if it results in improvement.

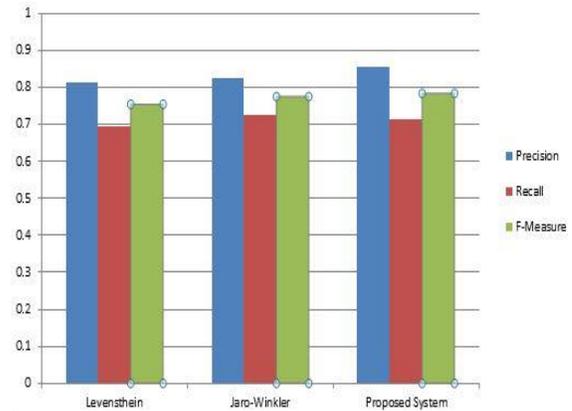


Fig. 2. Comparison of value with Dataset1.

3.3. Alignment Module

The alignment is normally calculated on the basis of two measures commonly known as precision and recall. The precision is the average of threshold with calculations of total number of classes and number of matching classes.

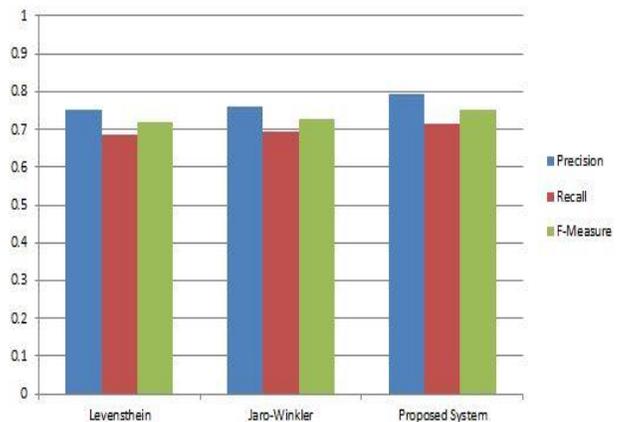


Fig. 3. Comparison of value with Dataset2.

$$\text{Precision, } P = \frac{|M_c * Avg_T|}{T_c} + \frac{|UM_c|}{T_c}$$

The recall is the average value of the predetermined values with some calculations described below.

$$\text{Recall, } R = \frac{|M_c * Avg_P|}{T_c} + \frac{|UM_c|}{T_c}$$

Where M_c is the number of matching class, Avg_T is the average of Threshold values, T_c is the total number of classes, UM_c is number of unmatched class and Avg_P is the average of prescribed value.

The F-measure (Final Measure) is the expected output result which is the average of Precision and Recall.

$$\text{Fmeasure, } F_m = \frac{P + R}{2}$$

For each string similarity metric, the alignment results are extracted for a certain threshold value ranging from 0.1 to 1. The alignment result consists of precision, recall and F-measure. In this evaluation, the string distance calculation like Levenstein, Jaro-winkler and SMOA are computed using the proposed algorithm. The computed results are shown in fig.2 and fig.3

The above figures show that proposed system is more efficient than the existing methods. Our system computes the alignment containing only correspondences between classes and properties which exclude individual matching. The existing methods take more computation time with less efficiency. The F-measure values are computed starting from average of precision and recall values.

5. CONCLUSION

Ontology Alignment is the key solution for Heterogeneity problem in semantic web. This solution will identify the correspondence between the entities of different ontologies. In this paper, we propose to use SMOA for calculating the correspondence of ontologies in order to replace previous string matching techniques. Our proposed system will categorize the ontology into class, subclass and various properties. It uses memetic algorithm with String matching technique. The experiment result shows that our approach provides optimal solution among previous methods. Comparing with state of the arts ontology matching system our proposed approach is able to get more accurate results. In future work, we want to extend this system with different types of matching strategies by taking further algorithms for efficiency and further application domain into account.

REFERENCES

- [1] P. Shvaiko, J. Euzenat, "Ontology matching: State of the art and future challenges", *IEEE Transactions on Knowledge and Data Engineering*, 25(1), 2013, pp.158-176.
- [2] W. Hu, Y. Qu, G. Cheng, "Matching large ontologies: A divide-and-conquer approach", *Data and Knowledge Engineering*, 67(1), 2008, pp.140-160.
- [3] G. Acampora, V. Loia, S. Salerno, & A. Vitiello, "A hybrid evolutionary approach for solving the ontology alignment problem", *International Journal of Intelligent Systems*, 27, 2012, pp.189-216.
- [4] Y. Sun, L. Maa, S. Wang, "A Comparative Evaluation of String Similarity Metrics for Ontology Alignment", *Journal of Information and Computational Science*, 12(3), 2015, pp.957-964.
- [5] X. Xue, Y. Wang, A. Ren, "Optimizing ontology alignment through Memetic Algorithm based on Partial Reference alignment", *Expert Systems with Applications*, 41, 2014, pp.3213-3222.
- [6] J. Euzenat, P. Shvaiko, *Ontology Matching*, 2nd edition. Springer-Verlag, Berlin Heidelberg, ISBN: 978-3-642-38721-0, 2013.
- [7] G. Stoilos, G. Stamou, & S. Kollias (2005), "A string metric for ontology alignment", In *Proceedings of 4th International Semantic Web Conference (ISWC 2005)*, pp. 623- 637.
- [8] OWLAPI, <http://owlapi.sourceforge.net/>.
- [9] INIS Thesaurus, <https://www.iaea.org/inis/links/index.html/>.
- [10] N.F. Noy, M.A. Musen, "PROMPT: algorithm and tool for automated ontology merging and alignment", *Seventeenth National Conference on Artificial Intelligence*, 2000, pp. 450-455.
- [11] J. Madhavan, P.A. Bernstein, E. Rahm, "Generic schema matching with cupid", *Proceedings of the 27th International Conference on Very Large Data Bases (VLDB)*, 2001.
- [12] J. Bock, J. Hettenhausen, *Discrete particle swarm optimisation for ontology alignment*, *Information Sciences* 192, 2012, pp.152-173.
- [13] Y.R. Jean-Mary, E.P. Shironoshita, M.R. Kabuka, "Ontology matching with semantic verification", *Journal of Web Semantics* 7 (3), 2009, pp.235-251.
- [14] J. Noessner, M. Niepert, "CODI: Combinatorial Optimization for Data Integration Results for OAEI 2010", *Proceedings of the 5th International Workshop on Ontology Matching*, 2010.
- [15] H.H. Do, E. Rahm, "Matching large schemas: approaches and evaluation", *Journal of Information System* 3 (6) (2007) 857-885.
- [16] G. Acampora, V. Loia & A. Vitiello, "Enhancing ontology alignment through a memetic aggregation of similarity measures", *Information Sciences*, 250, 2013, pp.1-20.
- [17] X. Xue, Y. Wang, *Optimizing ontology alignments through a Memetic Algorithm using both MatchFmeasure and Unanimous Improvement Ratio*, *Artificial Intelligence*, 223, 2015, pp.65-81.
- [18] F. Hamdi, B. Safar, C. Reynaud, H. Zargayouna, "Alignment-based partitioning of large-scale ontologies", *Advances in Knowledge Discovery and Management. SCI*, Springer, Heidelberg (2010), vol. 292, pp. 251-269.
- [19] Algergawy A, Massmann and Rahme, "A Clustering-Based Approach for Large-Scale Ontology Matching", *Springer-Verlag Berlin Heidelberg, ADBIS 2011*, pp. 415-428.
- [20] Abolhassani, H., Haeri, S. H. & Hariri, B., "Coincidence Based Scoring of mapping in Ontology Alignment", *Journal of Advanced Computational Intelligence and Intelligent Informatics* (2007), Vol.11 No.7, pp.1-14.